Police Officers Neglect Data Selection in Crime Predictions

Juan B. González Santiago De Martini Ervyn Norza Santiago M. Perez-Vincent *

November 30, 2025
(Most Recent Version Here)

PRELIMINARY AND INCOMPLETE DRAFT, PLEASE DO NOT DISTRIBUTE

Abstract

Crime data suffers from selection bias: only a subset of crime events becomes observable, and observability endogenously varies across locations and groups. If police officers form beliefs by taking selected data at face value rather than adjusting for the selection process, biased data can become biased crime predictions. We provide the first causal evidence on this mechanism using a lab-in-the-field experiment with 182 Captains of the National Police of Colombia, who routinely rely on reported crime data to allocate patrols. In a Crime Prediction Task, officers observe both crime reports and reporting rates and are incentivized to predict which neighborhood experienced more crime. Officers exhibit substantial selection neglect: accuracy falls by 36 percentage points when reports must be adjusted for reporting rates, and errors are largest when the true crime gap is wide. Neglect operates along two cognitive margins: one third of officers do not consider reporting rates at all (extensive margin), while many of those who do consider them apply the adjustment only when it is computationally easy (intensive margin). Selection neglect has first-order consequences: when reporting rates systematically differ across groups, officers systematically over-predict crime in the over-selected group, and endogenous feedback amplifies this pattern. A predictive policing algorithm that corrects for selection fails to improve predictions because officers who neglect selection override its corrective recommendations.

*González: Department of Economics, University of Southern California, juanbgon@usc.edu. De Martini: Department of Economics, University of Southern California, sd59576@usc.edu. Norza: National Police of Colombia, eh.norza10@uniandes.edu.co. Perez-Vincent: Inter American Development Bank, santiagoper@iadb.org. We thank Isabelle Brocas, Juan Carrillo, Giorgio Coricelli, Jonathan Libgober, Alejandro Martínez-Marquina, Paulina Oliva, Emanuel Vespa, and the participants of the Graduate Behavioral and Experimental Economics Conferences at Caltech and UCSB for thoughtful feedback on the study.

I Introduction

Police decisions, from patrolling to stops, searches, and arrests, are guided by officers' beliefs about where crime is most likely to occur and who is most likely to commit it. A large empirical literature documents persistent racial and socioeconomic disparities in these decisions (National Academies of Sciences, Engineering, and Medicine, 2023). Many mechanisms can generate such disparities, including taste-based bias, stereotypes, incentives, and deficient regulation. Recent evidence suggests that cognitive processes may also contribute to unequal policing outcomes (Dube et al., 2024; Ferrazares, 2025). Yet the formation of crime beliefs remains largely a black box, because administrative crime data cannot isolate what officers think from the data that they generate. Since crime data itself is produced by policing decisions, observational data cannot reveal whether officers correctly interpret crime data or whether their inference amplify or mitigate disparities.

A defining feature of crime data is that it is *selected*: only a subset of crime events are observed and recorded. Crucially, this selected subset might not be representative of the underlying universe. In many settings, observability varies systematically across groups and locations, creating selection bias in crime data. For example, frequent patrolling in minority neighborhoods increases the likelihood that crime occurring there is observed (Chen et al., 2023), and higher arrest rates for minority individuals generate over-representation in arrest datasets (Gonçalves et al., 2025). Therefore, a crime dataset that over-represents a given group does not necessarily mean that group commits more crime: it may mean that crime involving that group is more observable. A rational agent would adjust observed crime data using the data-generating process to infer underlying crime. But if decision makers take selected data at face value, biased data can translate into biased beliefs and, in turn, biased policing decisions. This mechanism does not require stereotypes, distrust, or animus: even unbiased algorithms fall into it when making crime predictions (Arnold et al., 2025). Although selection neglect is a common inference error (Brundage et al., 2024; Enke, 2020), this mechanism has proven difficult to test empirically in the context of policing because of the endogenous nature of crime data and policing decisions.

This paper provides the first causal evidence on whether police officers neglect the selection of crime data when forming beliefs about crime. We conduct a lab-in-the-field experiment with 182 Captains of the National Police of Colombia, the officers who use crime data to make patrol allocation decisions in the field. Colombia provides an ideal setting because crime data is selected primarily through citizen reporting, making the data-generating process highly transparent: the reporting behavior directly maps into the probability that a crime becomes observable. Unlike settings where selection is produced by police enforcement (e.g., stops or arrests), officers in Colombia have no strategic incentives to ignore selection in reported crime data, allowing us to cleanly isolate the cognitive dimension of neglect. Our design mirrors the core inference problem faced by officers in the field: crime is selectively observed, the main source of crime data is citizen reports, and reporting rates vary across neighborhoods. In the Crime Prediction Task, officers see the number of crime reports received in each of two *fictitious* neighborhoods, and are incentivized to correctly guess in which neighborhood of the pair more crimes actually happened based on this selected data. Importantly, officers know the data generating process and its selection: they observe the probability that a crime is reported in each neighborhood, that we call their reporting rate. By randomizing within-subjects the reporting rates across neighborhoods, we generate exogenous variation in selection bias and observe whether officers adjust for or neglect it [1]. The design isolates the inference problem while abstracting from taste-based motives.

We document three sets of results. First, we find robust evidence of selection neglect in crime predictions. The exogenous variation in selection bias produces two cases: decisions where the neighborhood with more reports is also the neighborhood with more crime, so the naïve and the selection-adjusted predictions are identical, and others where adjusting for selection is needed. By comparing prediction accuracy across those decisions, we are able to identify selection neglect net of other inference errors. This approach shows that officers make the correct prediction in 77% of decisions where selection bias doesn't change the optimal guess, but only in 41% of decisions where adjusting for selection is necessary: an accuracy loss of 36 percentage points. These errors are not confined to "close calls": selection neglect is most pronounced precisely when the crime gap between neighborhoods is large, meaning errors are most damaging in high-stakes predictions. As a result, officers neglect in their predictions an additional 6.4 crimes due to selection, equivalent to 40% of the average crime gap between neighborhoods.

Second, we identify two complementary mechanisms behind selection neglect. On the extensive margin, 37% of officers choose not to view the reporting rate even when doing so is costless, revealing that selection does not enter their mental model of crime data. Selection doesn't fail to come to mind due to perceptual salience: making reporting rates more salient in later decisions does not reduce neglect among officers who initially failed to consider data selection. Instead, neglect is predicted by officers' beliefs about how reliable the reported crime data they work with in the field is: officers who believe reported crime data to be a biased mapping of real crime are better at correcting for selection. On the intensive margin, officers who do consider reporting rates frequently fail to apply the adjustment because doing so is computationally costly: correcting for selection takes longer response times, but is more likely when reporting rates are simple multiples. Once officers both recognize and compute the adjustment, e.g., when using the on-screen calculator, neglect disappears entirely. The dual-margin mechanism is further supported by the bimodal distribution of accuracy: 30% of the sample never corrects for selection, 20% always do, and half of the sample adjusts selectively.

Third, we document that selection neglect has first-order consequences for crime beliefs: it directly maps selection bias into biased predictions. When reporting rates systematically differ across groups of neighborhoods, officers systematically over-predict crime in the over-selected group despite identical true crime levels. When selection is made endogenous, selection neglect creates cycles of bias: officers predict more crime in over-selected areas, increase observability there, and reinforce the selection bias. Under combined systematic and endogenous selection, 37% of officers lock into a self-reinforcing pattern of overprediction. These disparities emerge despite the complete absence of stereotypes, taste-based motives, or operational incentives, isolating a purely cognitive channel through which biased data becomes biased beliefs.

Finally, we test potential interventions to debias crime predictions. We first evaluate predictive policing algorithms, a widely proposed solution to human inference biases. We compare an algorithm

[1] For example, take the case where neighborhood A received 120 crime reports, and neighborhood B received 100. A neglecting officer would take the crime reports at face value and predict that more crimes happened at neighborhood A. An officer that accounts for selection would pay attention to the reporting rates, 75% and 50% respectively, and predict that more crimes happened in B.

that corrects for data selection to one that takes reports at face value. On average, neither improves crime predictions. The null effect is the result of a striking pattern: officers who already adjust for selection follow the correcting algorithm and override the neglecting one, whereas officers who neglect selection override the correcting algorithm and follow the neglecting one. In other words, the very officers the algorithm is meant to help reject corrective recommendations when they contradict their heuristic. This finding complements recent evidence that judges override risk algorithms in ways that reduce decision quality (Angelova et al., 2025).

This paper contributes first and foremost to a growing experimental and theoretical literature on how individuals learn from data under misspecified representations (Bohren & Hauser, 2025; Bohren et al., 2024; Bordalo et al., 2023). A series of laboratory studies has documented selection neglect, the tendency to treat selected datasets as if they were random samples when updating beliefs (Ali et al., 2021; Barron et al., 2024; Enke, 2020; Esponda & Vespa, 2018; Farina & Herman, 2025; Jin et al., 2021; Koehler & Mercer, 2009). We extend this literature in two directions. Empirically, we show that selection neglect is pervasive among trained decision-makers operating in the field, not only among experimental subjects completing abstract tasks. We shed light on why neglect occurs by separating two cognitive margins: whether selection enters the agent's mental model at all, and whether it is correctly adjusted once considered. Conceptually, our setting demonstrates not only that neglect persists among experts, but also its consequences: neglect distorts beliefs precisely when those beliefs guide high-stakes decisions, creating systematic disparities in predictions. Thus, the main contribution of the paper is to identify the extent, mechanisms, and belief-level consequences of selection neglect among police officers: the population for whom this inference problem matters in practice.

Second, the paper speaks to an active literature on racial and class disparities in policing (National Academies of Sciences, Engineering, and Medicine, 2023). Prior work has established disparities in enforcement outcomes (Aggarwal et al., 2025; Chen et al., 2023; Fryer, 2019; Gonçalves et al., 2025; Pierson et al., 2020), highlighting heterogeneity across officers (Ba et al., 2021, 2025; Goncalves & Mello, 2021; Hoekstra & Sloan, 2022; Rozema & Schanzenbach, 2019; Vomfell & Stewart, 2021), and analyzing institutional drivers (Ferrazares, 2024; Rivera, 2025; Rozema & Schanzenbach, 2023; Stashko, 2023). Recent work has turned toward the cognitive and emotional processes shaping enforcement (Feigenberg & Miller, 2025; Ferrazares, 2025; Holz et al., 2023), motivated by the idea that improving decision processes may be more feasible than changing preferences. For instance, Dube et al. (2024) and Owens et al. (2018) find that training officers to use more reflective thinking reduces use of force and enforcement disparities. We complement this research by focusing on a precursor to enforcement: belief formation. In our setting, biased crime predictions emerge solely from a failure to account for selection in crime data, without stereotypes, taste-based motives, or institutional incentives. Our results therefore do not claim that biased beliefs are the primary driver of enforcement disparities, nor that discrimination in the field is reducible to cognitive error. Instead, we show that when crime data are selected, selection neglect provides a behavioral pathway through which biased data generates biased beliefs, potentially feeding into downstream policing decisions. In doing so, the paper bridges behavioral economics and the policing literature by identifying a cognitive mechanism that shapes the beliefs officers bring to the field. While different departments and officers might have different objective functions when making policing decisions, unbiased predictions are needed for efficient and fair enforcement.

The paper proceeds as follows. Section II introduces the setting of policing in Colombia, and our sample of police officers. Section III describes the experimental design tailored for our field setting, and the empirical strategy. Section IV presents the results on the extent and mechanisms of selection neglect in crime predictions, and Section V documents the consequences of this bias. Section VI presents the results on some interventions designed to mitigate neglect. Section VII discusses related literature and concludes.

II Setting

The experiment is a stylized implementation of the inference problem that police officers face when using crime reports to make enforcement decisions. Rather than attempting to reproduce the full complexity of policing decisions, the design isolates the belief-formation component of that problem: given the distribution of reported crime data, where does the officer believe more crimes happened? Throughout the paper, a neighborhood choice or guess is therefore interpreted as a discrete crime prediction, not as an equivalent to patrolling or policing decisions. The core identification challenge in the field is that crime is latent and crime data is selected: only a subset of crimes are reported, and reporting rates differ across places and over time. If officers adjust for how crime is selected into the dataset, their predictions will track true crime; if they neglect selection, their predictions will track reporting patterns instead. The experiment replicates this statistical structure in a fully transparent environment, allowing us to measure whether officers adjust for or neglect selection when generating crime predictions. Importantly, participants are Captains of the National Police of Colombia: precisely the officers who routinely interpret crime data and make patrol-allocation decisions in the field. This ensures that the experiment measures the cognition of the real decision makers responsible for police deployment. The rest of the section describe the field setting, and our sample.

A Data-Driven Policing in Colombia

The National Police of Colombia uses crime data to guide their patrolling assignments and resource allocation. Because we design our experiment to mirror the inference problem officers routinely face in the field, we highlight here some of the main features of data-driven policing in Colombia.

1. *Crime predictions guide policing decisions.* Station commanders infer the geographic and temporal distribution of crime within their attention zone (equivalent to *beats* in the United States), and take these trends as inputs in patrol allocation decisions. While officers might have different goals when assigning patrols, forming accurate beliefs is crucial for any objective, no matter whether it is maximizing arrests or minimizing crime.

2. *Crime reports are the main source of information used for crime predictions.* Officers are trained to use *reported crime data* as the main input of their predictions. Station commanders routinely receive briefs with the geographic distribution of crime reports within their zone of attention. Importantly, these briefs aggregate all types of crime into a single metric: number of reports.

3. *Reported crime data suffers from selection bias.* According to national victimization surveys, only 29% of crimes in Colombia are ever reported to the police (DANE, 2023). The crimes that are observed and reported are not a representative sample of the universe of crimes. There are systematic

differences in which types of crime are observed —while most homicides are uncovered, only 7% of extortion cases are ever reported, and where —while 45% of crimes are reported in Bogotá, only 16% are in rural areas. While there is little causal evidence on the effect of policing on reporting behavior in Colombia[2], mistrust in the police is cited as one of the main reasons for not reporting a crime.

These institutional features motivate the core elements of our experiment. Officers in Colombia must infer crime levels from selected data under time pressure and uncertainty. If officers neglect the selection of reported data, their crime predictions might be driven by reporting and not by actual crime, leading to inefficiencies and policing disparities. Our sample comprises precisely the officers that use selected crime reports to predict crime and assign patrols, allowing us to identify whether they neglect or correct selection bias in a controlled environment that preserves the essential structure of their real inference problem.

It's important to note that, although reporting rates is the main source of selection of crime data in this setting, it's by no means the only source of data selection in other policing contexts. In the US, police agencies use police generated data to guide policing strategies and decisions: stops, arrests, police-generated reports, etc (Brayne, 2020). All these data sources have a similar selection problem: not all suspects are stopped or arrested, not all crimes are observed, but only a *selected* subset. Any selection bias in which crime events are selected into crime data will map into over- and under-representation of certain places and groups in crime data. For instance, because Black people are more likely to be arrested, they are over-represented in arrest data, compared to the potential arrestee population (Gonçalves et al., 2025). Similarly, because police officers patrol more frequently around minority neighborhoods, crime events are more likely to be observed and thus enter crime data in those areas (Chen et al., 2023). In conclusion, while reporting rates are the main selection mechanism in this setting, selection bias in crime data is a pervasive feature of policing more generally.

B Sample

Logistics — In September 2025, 202 Captains of the National Police of Colombia were in the Police Postgraduate School (ESPOL) in Bogotá, receiving the training required to promote to Majors (NATO OF-4). These 202 officers comprise the universe of promoting Majors in Colombia, and were instructed by superior officers to be present at our experimental session. Once checked in the session, participants were invited to participate in a "Police Science Study", and offered an incentive of 5 to 20 USD. 18 officers (8.9%) decided not to participate at this stage, before knowing anything about the experiment. Of the 184 officers who started the experiment, all but two (99%) completed it, so there are no concerns about endogenous attrition.

Final Sample — 182 Captains complete our lab-in-the-field experiment. We target Captains because station commanders, who are in charge of assigning patrolling officers over time and space, are usually captains. Therefore, our sample has relevant experience making patrolling decisions in the field. In particular, 25% of our sample was acting as station commander just before starting training, and another 40% was tasked with operative decisions involving allocating police resources. Officers in this sample have

[2]See Ang et al. (2025) for causal evidence in the US, and Perez-Vincent et al. (2024) for a review of evidence from Latin America.

Table 1: Mapping Field Features to Experimental Design

Field Feature (National Police of Colombia)	Experimental Analogue
Station commanders make crime predictions to guide patrolling policing decisions.	Officers are incentivized to guess the neighborhood with higher-crime of a pair.
Crime predictions rely primarily on reported crime data.	Participants observe reported crimes r_{gn} and must infer latent crime levels c_{gn}.
Selection into reported data varies systematically across areas (reporting gaps).	Reporting rates s_{gn} vary across neighborhoods, generating selection bias.
Systematic selection bias across groups can generate policing disparities.	In the *Systematic Selection Bias* treatment, one city has higher reporting rates on average.
Crime predictions affect future data collection (endogenous selection).	In the endogenous selection treatment, patrolling a city increases its reporting rate in the next round.
Station commanders are trained to use crime predictions to guide policing decisions.	Our sample comprises station commanders and operative officers with direct responsibilities.

Features to isolate selection neglect in inference:

– Crime is generated from city-specific distributions, mirroring crime differences across groups but ruling out taste-based discrimination.

– Reports equal expected values ($r_{gn} = s_{gn}c_{gn}$), removing sampling noise to isolate inference errors cleanly.

– Officers predict crime across two neighborhoods rather than over an entire grid.

an average of 14 and a minimum of 12 years of policing experience, have at least an undergraduate degree —a requirement for being promoted to this rank, and 15% of them have a masters degree. Additionally, because our sample of Captains is that of promoting Captains, if anything it is positively selected in terms of ability. In terms of demographics, 83% of these officers are men, and range in age from 32 to 45 years old (average 36 years old). Because our sample consists of Captains promoting to Majors, these officers have deep experience with patrolling decisions and are the very individuals responsible for interpreting crime data in the field.

III Design

Our experimental design is a stylized version of the crime inference problem station commanders face when allocating patrols and making decisions on the field. The task strips away operational complexities while preserving the core features of their prediction environment. First, participants must infer crime to guide policing decisions. Because officers may have diverse objectives when making these decisions, we abstract from outcomes such as arrests or deterrence and focus on the crime inference problem. Since officers are used to making implicit and relative crime predictions rather than explicit ones, we frame the task as a sequence of binary choices instead of eliciting continuous beliefs. This ensures that participants experience the task as a mirror of the inference problem they routinely face in the field, rather than as an abstract math exercise. Second, our design presents reported crime data as the main input to guide crime predictions. Third, reported crime data suffers from selection bias, which we summarize in a single, transparent metric: the reporting rate. By simplifying the environment, we can precisely identify whether and how officers correctly adjust for the selection of reported crime data, that is, whether they exhibit selection neglect when making crime predictions.

A Crime Prediction Task

We now describe the Crime Prediction Task, the core decision environment officers face in the experiment. Each element of the task corresponds to a specific field feature outlined in Table 1. Participants see fictitious data for two cities g, city A and city B, each consisting of a continuum of neighborhoods. In each neighborhood gn a number of crimes c_{gn} occurs, which is unobservable to the participant.[3] In each round, officers see a pair of neighborhoods An, Bn and are incentivized to predict the neighborhood with the highest number of crime. In particular, participants are paid \$5 if they guess the correct neighborhood in one randomly selected decision. Binary predictions allow us to cleanly detect whether officers adjust reported data by its reporting rate or instead rely on reports at face value, a signature of selection neglect, without imposing an elicitation method that officers might find unnatural.

To predict the crime level c_{gn}, participants can use crime data. Participants see the number of crimes reported r_{gn}, as well as the reporting rate s_{gn} of each neighborhood, i.e., the probability that a crime is reported. Because the observability of crimes is determined by reporting, reporting rates determine the selection of crime data. This way, two neighborhoods with the same number of crimes can have different numbers of reports just because of differing reporting rates. Additionally, in the first part of the experiment, participants can use demographic data $\mathbf{X}_{gn}$: the average age, unemployment rate, and the average socio-economic status (SES).[4] Demographic data in this setting serves two purposes. First, these demographic variables act as a decoy that reduces the salience of data selection. Second, if officers still use this data to guide their decisions in a setting where crime data is perfectly predictive, it provides a measure of statistical discrimination in their routine decisions.

Data Generating Process (DGP) — The crime level of each neighborhood is an independent draw from a city-level crime distribution, $c_{gn} \sim \mathcal{N}(\mu_g, \sigma^2)$. Importantly, officers have no prior information about the parameters of these distributions. The reporting rate s_{gn} is drawn independently from a uniform distribution $\mathcal{U}[0.1, 0.9]$ and rounded to the closest multiple of 5 to ease calculations. We calculate the number of reported crimes based on the number of crimes and the reporting rate. In theory, because each crime that occurred in neighborhood gn can either be reported or not, and the report probability is s_{gn}, the number of crime reports can be thought of as a binomial random variable $r_{gn} \sim Bin(c_{gn}, s_{gn})$, with $E[r_{gn}] = c_{gn}s_{gn}$. Therefore, the number of reports is *selected* through the reporting rate, and to infer the number of crimes from reported data the optimal prediction is to *adjust* the reports by the reporting rate: $\hat{c}_{gn} = r_{gn}/s_{gn}$. To reduce noise in our data generating process, we randomly draw c_{gn} and s_{gn} and then fix the number of reports to its expected value: $r_{gn} = c_{gn}s_{gn}$. This ensures that a decision maker who correctly adjusts for selection always makes the correct guess, so that any deviation from the optimal prediction must come from cognitive error and not from sampling noise. Finally, demographic variables $\mathbf{X}_{gn}$ are independently drawn from uniform distributions, and thus are completely unpredictive of crime.

What participants know about the DGP — The structure of the DGP mirrors the core statistical challenge officers face in the field, inferring latent crime from incomplete data, while remaining simple

[3] Cities can be understood as a placeholder for groups which might be perceived to have different distributions of crime: race, SES, etc. By using abstract cities instead of any of these groups, we strip away any taste-based discrimination from this setting.

[4] In Colombia, residential buildings are classified according to their socio-economic status (*estrato*), in a scale from 1 (low) to 6 (high). The *estrato* is used to determine utilities prices, and to target public services, so it is a very familiar proxy of income for officers.

enough to be explainable without statistical training. Participants know that each neighborhood has an uncertain level of crime c_{gn}, but they do not know the primitives of the city-level distributions. Officers also learn that the reporting rate s_{gn} is the probability that a crime is reported, so that in expectation $E[r_{gn}] = c_{gn}s_{gn}$.[5] To make sure participants understand how to adjust for selection, they answer comprehension questions such as: "*In an area with 50% reporting rate, 40 crime reports were received. What is the best guess of how many crimes happened in that neighborhood?*" Finally, participants can have any belief about the relationship between demographics and crime, but they know that they can infer the number of crimes just from the number of reports and the reporting rate.

B Experiment Structure

The experiment is divided in two main parts, preceded by instructions and a short training, and followed by a short survey. Part I sets expectations and allows officers to become familiar with the interface. Part II allows a more clean identification of selection neglect, in a setting where selection is made even more salient.

Training Decisions — After listening to the instructions, participants complete 5 training decisions so they become familiar with the interface. They have the opportunity to ask any questions after this initial exposure to the task.

Instruction Comprehension Questions — To ensure that they have understood the instructions, participants respond to 9 questions testing their comprehension of the task and the data generating process. Wrong answers are corrected.

Math Questions — Because basic math operations are needed to correctly adjust for selection in our setting (i.e., dividing the number of reports by the reporting rate), we test these skills through two incentivized questions ($2 each).

Part I — In Part I, officers complete 10 crime predictions across neighborhoods of cities A and B. In the first decision, officers only see the number of reports for each neighborhood. We use this choice as a sanity check: officers should patrol the neighborhood where more reports were received. Then, for the next 8 decisions, officers see all crime and demographic information. For the tenth decision, participants can see the number of reports plus one other variable of their choice. So, before seeing any data of the tenth pair of neighborhoods, officers decide which variable they consider more relevant for their decision, which will be shown along the number of reports. This choice gives insight about the mental model of officers. If they conceptually understand selection is relevant for crime prediction, they should choose to see the reporting rates. Choosing any other variable is indicative of selection neglect.

Participants have 1 minute to make each prediction in Part I. To ease computations, officers can use a calculator on screen. In 3 decisions of Part I, we elicit subjects' cognitive uncertainty, this is, their uncertainty about whether they had made the optimal guess, using the wording of Enke and Graeber (2023). We randomly select one decision of Part I for payment, and pay $5 if officers chose the neighborhood with higher number of crimes.

[5] In the instructions, we give examples such as: "*Imagine that 200 crimes happened in a neighborhood. If the reporting rate is 50%, we can expect 100 crimes to be reported. Thus, we can expect 100 crime reports from that area*".

Part II — In Part II, officers make 12 crime predictions across pairs neighborhoods, each from a different city, C and D. Throughout these 12 decisions, participants see only the number of reports and the reporting rate of each neighborhood, but no demographic data. Because the only available information is reports and their selection, the selection problem is more salient. As in Part I, participants have 1 minute to make each decision and can use an on-screen calculator. We elicit their cognitive uncertainty twice through this part. We randomly select one decision of Part II for payment, and pay $5 if officers patrolled the neighborhood with higher number of crimes.

To test whether selection bias across groups can generate statistical discrimination, we randomly assign officers into two treatments for Part II. For those in the *Systematic Selection Bias* treatment, one city has systematically higher reporting rates than the other: $s_{Cn} \sim \mathcal{U}[0.1, 0.6]$ and $s_{Dn} \sim \mathcal{U}[0.4, 0.9]$ or vice versa. This creates a 30pp reporting gap between cities in expectation. This treatment mirrors real crime data, where some groups have systematically higher likelihood of being selected into crime data (Gonçalves et al., 2025). Reporting rates in the *Uniform Selection Bias* treatment are independently drawn from the same distribution for both cities, as in Part I.

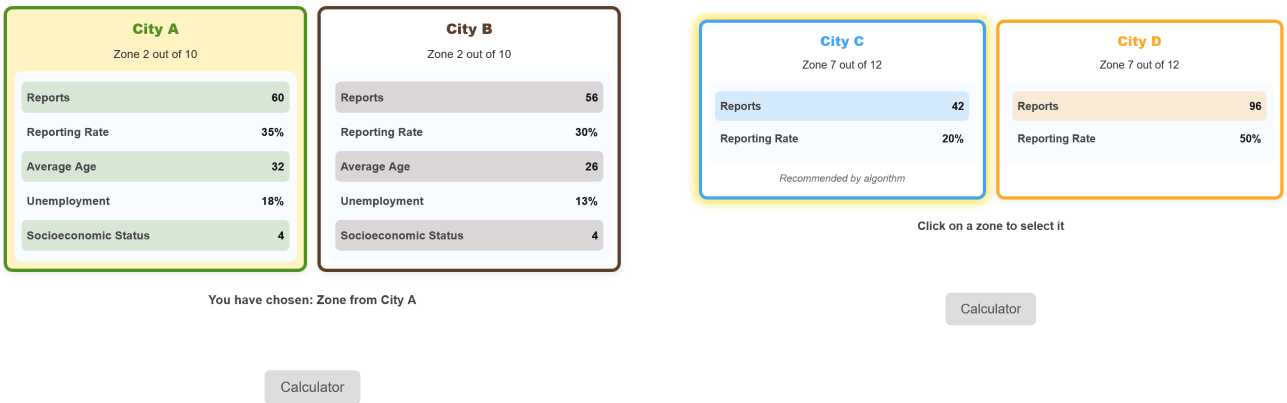
To explore whether endogenous data selection reinforces disparities under selection neglect, we randomly assign officers to two treatments only for the last six decisions of Part II. Selection remains exogenously determined in the *Exogenous Selection* treatment, drawn from the distributions specified above and unaffected by patrolling choices. In the *Endogenous Selection* treatment, the past crime prediction affects data selection. In particular, if the officer chooses city g in decision $n-1$, we add a "reporting bonus" of 15 percentage points to that same city in the next decision: $s'_{gn} = s_{gn} + 0.15$. This treatment mirrors endogenous selection of crime data outside the lab, as policing decisions affect data selection through increased crime observability and changes in citizen reporting (Ang et al., 2025; Chen et al., 2023).

Finally, we test a common intervention in policing that aims to improve crime inference: predictive policing algorithms, which have become pervasive across police departments (Brayne, 2020). We implement two algorithms. Both take as input the reported crime data on the choice neighborhoods. From reported crime data, the algorithms estimate which neighborhood is likely to have more crimes and issue a patrol recommendation for that neighborhood, with some noise. The key difference is that one algorithm, which we label the *Correcting Algorithm*, adjusts for data selection (reporting rates) when making this prediction, while the *Neglecting Algorithm* only uses raw reports. Participants are informed of whether their algorithm corrects for selection or not. This allows us to test whether selection neglect interacts with algorithmic recommendations, i.e., whether algorithms can compensate for selection neglect or whether neglect extends to ignoring the algorithm. We randomize participants into three conditions for the last 6 decisions of Part II: control (no algorithm), correcting algorithm, or neglecting algorithm.

End Questions — After completing the experiment, participants complete a short unincentivized survey. We ask them to guess how many correct predictions they made through the experiment, as well as whether they would like to receive feedback on how to improve their crime predictions. Most importantly, officers self-report their beliefs on whether reported crime data measures actual crime with or without bias, across several dimensions: whether crime reports identify temporal and spatial trends in crime, whether crime reports are the best available data, whether citizens know how to report a crime, and whether reported data is biased. These questions allow us to connect experimental inference errors to

Figure 1: Crime Prediction Task Screens

(a) Part I - Low Selection Salience **(b)** Part II - High Salience and Algorithm



Notes: Panel (a) shows the decision screen for Part I, in an example where the participant has guessed the neighborhood from City A to have more crime. Note that in Part I, reports and reporting rates are accompanied by demographic decoys, and that they are always presented as the first data sources. This makes it more difficult to to neglect selection. Panel (b) shows the decision screen for Part II, in an example where the participant has not yet made a prediction but sees the recommendation of the algorithm as a golden halo around the zone and a small span in the bottom. Algorithm predictions, as detailed below, are only provided for the last six decisions of Part II and for a subset of participants. Without the recommendation, the screen is identical but without the halo and the text at bottom. Note the on-screen calculator available throughout the experiment.

officers' real-world beliefs about biases in crime data by exploring heterogeneity in selection neglect along this dimension. Then, officers report their experience (in years) in the National Police of Colombia, as well as whether their last position before training involved administrative, operative, or surveillance tasks. Captains conducting surveillance tasks are those acting as station commanders and personally assigning patrols. Finally, officers answer basic demographic questions.

Table 2: Summary of Experimental Design

	10 patrolling decisions
	First decision: only see number of reports, sanity check
Part I	8 decisions: see number of reports, reporting rate, demographics
	Tenth decision: see number of reports and 1 data of their choice
	12 patrolling decisions
Part II	All decisions: only see number of reports, reporting rate
	Last 6 decisions: two selection treatments and algorithm interventions
End	Performance guess, trust in reported data, experience, and demographics

C Empirical Strategy

Our analysis follows directly from the experimental design. For each officer i and decision n, we observe which neighborhood gn they predict to have higher crime. For every decision, there is an unique *accurate* prediction: choosing the neighborhood with the highest number crimes c_{gn}. Although officers don't directly observe c_{gn}, the optimal choice rule using adjusted reports $\max_g r_{gn}/s_{gn}$ is always accurate, because we removed sampling noise from reported crime data. The optimal prediction does not depend on any priors or assumptions about the underlying distributions. Throughout the analysis the officer–decision pair (i, n) is the unit of observation. Because our prediction task uses binary choice, we use the *gaps* between neighborhoods $\Delta(x_n) \equiv x_{An} - x_{Bn}$ as the main decision features

Our empirical strategy focuses on whether officers adjust for or neglect data selection when making crime predictions. In our setting, neglecting selection corresponds to taking reported crimes at face value, i.e., ranking neighborhoods by r_{gn} rather than by r_{gn}/s_{gn}. However, because reporting rates are independently drawn for each neighborhood, neglecting selection has no effect on the ranking in some decisions. In these *selection-neutral decisions*, the neighborhood with more reports is also the neighborhood with more crimes, so that $\Delta(r_n)\Delta(r_n/s_n) > 0$. Selection neglect is irrelevant in these cases: even an officer who ignores reporting rates makes the correct prediction. We therefore use selection-neutral decisions as a benchmark that absorbs inference errors unrelated to neglect, such as prior beliefs or misunderstandings of the task.

In contrast, there are decisions where adjusting for selection reverses the ranking of neighborhoods. In these *selection-sensitive decisions*, the neighborhood with more reports is *not* the one with more crimes, so $\Delta(r_n)\Delta(r_n/s_n) < 0$. In these cases, neglecting data selection mechanically pushes officers into the wrong choice. It's important to note that selection-neutral and selection-sensitive decisions are almost identical apart from the fact that, in the latter, neglecting selection leads to the wrong prediction[6]. The difference in prediction accuracy between selection-neutral and selection-sensitive decisions thus captures the prediction error attributable to selection neglect.

Although accuracy is a natural measure of inference quality in this environment, it abstracts from the relevance of officers' mistakes. If officers were only neglecting selection in cases where the underlying crime gap is minimal, a large accuracy loss would not necessarily imply large policing errors. To assess the relevance and not only the prevalence of selection neglect, we define *neglected crimes* as the difference in crimes between the optimal guess and the actually chosen neighborhood. This metric captures the number of crimes that would have been guessed under the optimal rule but are neglected in the officer's prediction.

IV Selection Neglect in Crime Predictions

Our lab evidence from decision makers in the field shows that selection neglect in crime prediction is large, robust, and economically meaningful. Officers are highly accurate when selection does not matter, but accuracy drops sharply when selection must be corrected. This loss is not due to confusion or lack of math skills. Instead, two mechanisms operate in tandem: many officers do not incorporate reporting rates into their prediction process, and many others do incorporate them but fail to execute the adjustment when computations are effortful. These frictions are related to officers' experience with crime predictions and

[6] For instance, examine the case when $r_A = 30, r_B = 60$. If $s_A = 21\%, s_B = 40\%$, adjusting for or neglecting selection doesn't change the resulting choice: both the raw reports and the selection-adjusted reports point to B having more crimes, so this would be a *selection-neutral decision.* If $s_A = 19\%, s_B = 40\%$, neglecting selection leads to the wrong choice: raw reports point to B as the more criminal neighborhood, but adjusting for selection reveals it's A the area with more crimes. This last case would be a *selection-neutral decision.* This example illustrates that both types of decisions look similar, and that the difference in accuracy between them can only be attributed to selection neglect.

their beliefs that real reported crime data is a reliable reflection of underlying crime. Once both frictions are removed, when officers understand that selection must be corrected and can compute the correction easily, selection neglect disappears.

A The Accuracy Cost of Selection Neglect

We begin by quantifying how much selection neglect reduces the accuracy of crime predictions. In decisions where neglecting selection does not change the optimal prediction (*selection-neutral* decisions), officers choose the correct neighborhood in 78% of cases. In contrast, accuracy falls to 42% when the correct prediction requires adjusting for selection (*selection-sensitive* decisions). When regressing accuracy on whether selection affects the optimal prediction, Table A.2 documents that selection neglect causes a 35.8 pp ($p < 0.001$) loss in prediction accuracy. This loss is not only statistically significant but consequential: because the optimal prediction is to target the neighborhood with higher crime, errors translate directly into neglected crimes. Officers miss an average of 11.3 more crimes in selection-sensitive decisions ($p < 0.001$), a 114% increase relative to selection-neutral decisions and equivalent to roughly one quarter of the average crime gap between neighborhoods in Part I. Selection neglect therefore meaningfully hinders the effectiveness of crime prediction.

Panel (b) of Figure 2 finds that this accuracy loss is not explained by confusion or limited math skills. Officers who answered all nine instruction comprehension questions correctly display the same level of selection neglect as those who made at least one error ($p = 0.984$). Officers who completed all math questions with perfect accuracy also suffer a large and significant accuracy loss (29 pp, $p < 0.001$). Although neglect is slightly more pronounced among officers who made a math mistake ($p = 0.055$), the effect is substantial across both groups. In short, selection neglect arises despite understanding the task and having the basic arithmetic needed to solve it.

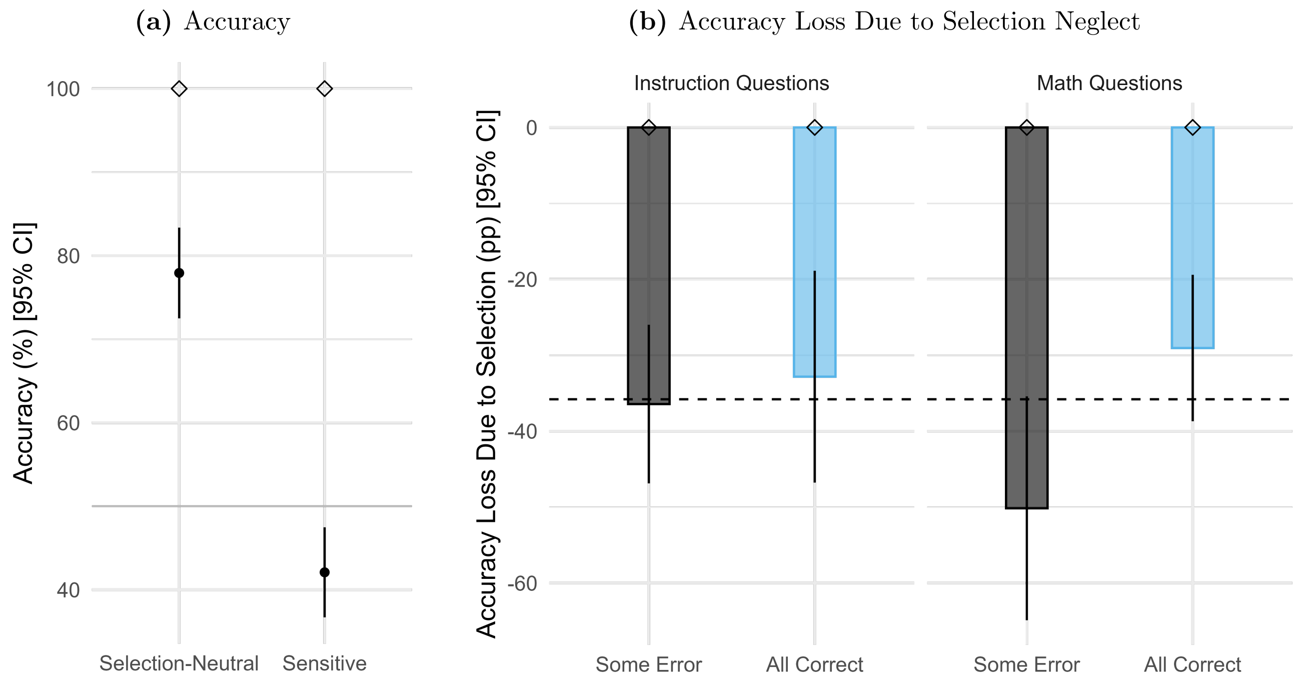
The evidence so far shows that selection neglect is both large and costly. We now ask why officers neglect selection: does selection fail to enter the prediction process in the first place, or does the adjustment break down once it comes to mind?

B Framework

We now delineate a simple conceptual framework to build intuition about the possible mechanisms of selection neglect. To correctly adjust for selection in our setting, the decision maker has to go through two steps of reasoning. First, selection needs to come to mind, this is, the decision maker needs to pay attention to selection as a feature of the statistical problem. Second, the decision maker needs to apply the correct adjustment rule to correct for selection bias, conditional on paying attention to selection. Because of the sequential nature of these two processes, we refer to them as the *extensive* and the *intensive margins* of neglect, respectively.

The inference problem of crime prediction can be described by a few features: the number of reports received in each neighborhood, and the selection of this data captured by the reporting rates. At the moment of inference, the decision maker chooses to which features of the problem to pay attention to. If the agent doesn't pay attention to data selection at this stage, we denote it as selection neglect at the extensive margin. Several factors can explain neglect at this stage. First, a growing literature shows how bottom-up salience can drive attention to some problem features and not others (Bordalo et al.,

Figure 2: Accuracy Loss Due to Selection Neglect



Notes: Panel (a) shows the average prediction accuracy in selection-neutral decisions (left column) and selection-sensitive decisions (right column). The panel documents an average accuracy loss of 35.8pp. Error bars show the 95% confidence interval of the mean, clustered at the officer level. Diamonds show the "Bayesian" benchmark, and the gray line shows the prediction for a random responder. Panel (b) shows the accuracy loss due to selection, calculated as the marginal effect of selection-sensitivity on accuracy. Error bars show the 95% confidence interval of the marginal effect. Results are disaggregated across sample subsets to show robustness. The left sub-panel divides the sample between those participants who made some error in the 9 instruction comprehension questions (68% of the sample, in black), and those who answered all 9 questions correctly (32%, in blue). The right sub-panel divides the sample between those who made some error in the 2 math questions (33% of the sample, in black), and those who answered all 2 questions correctly (67%, in blue). The dashed horizontal line marks the average marginal effect for the whole sample (35.8 pp).

2022, 2023). Second, mental models determine what features of the problem are attended, so a mental model that doesn't integrate selection will result in neglect (Jehiel, 2018; Schwartzstein, 2014). Our experiment is designed to provide evidence of neglect at the extensive margin through several ways. We make subjects choose what features of the problem they want to see for one prediction, revealing which feature they pay most attention to. Additionally, we exogenously vary the number of features by showing decoy demographic data in Part I but not in Part II. If selection neglect is exclusively driven by the extensive margin, we should find no neglect among those participants who choose to view selection, and it should decrease when increasing the salience of this feature in Part II. Previous experiments point to nudges and salience as effective tools to debias participants (Enke, 2020; López-Pérez et al., 2022).

Conditional on selection coming to mind, the correct selection-adjustment rule has to be applied. In our problem, the adjustment rule is intentionally simple: the decision maker only needs to divide the number of reports by the reporting rate (r/s) to get the best guess of crimes. However, there are several reasons why a decision maker can fail to adjust. Despite knowing selection is relevant, they might hold a conceptual misunderstanding on how it affects the data generating process (Jin et al., 2021; Koehler & Mercer, 2009). For instance, Farina and Herman (2025) and Esponda and Vespa (2018) find less neglect when selection is more predictable and understandable. Additionally, officers might know how to adjust but find it too cognitively taxing, selectively resorting thus to simpler rules —like taking the number of reports at face value— when decisions become more complex (Arrieta & Nielsen, 2024). Our design allows to explore the intensive margin of neglect in a variety of ways. If a deeper conceptual understanding of selected data generating processes matters, officers who are more aware of the biases of reported data should be better at correcting for selection. The distribution of average accuracy at the officer level provides evidence on the stability of adjustment: some officers might fail or succeed to adjust consistently, while others do it selectively. To shed light on how complexity leads to occasional adjustment failures, we exploit variation in how easy the adjustment computation is: when rates are multiples of each other, the computation becomes almost trivial. Finally, non-choice data, like response times and calculator use, also provide valuable information about the process of adjusting for selection bias.

C Evidence on Two Margins of Selection Neglect

Officers do not, on average, adjust sufficiently for selection bias when forming crime predictions based on reported crime data. We now turn to the mechanism behind this result.

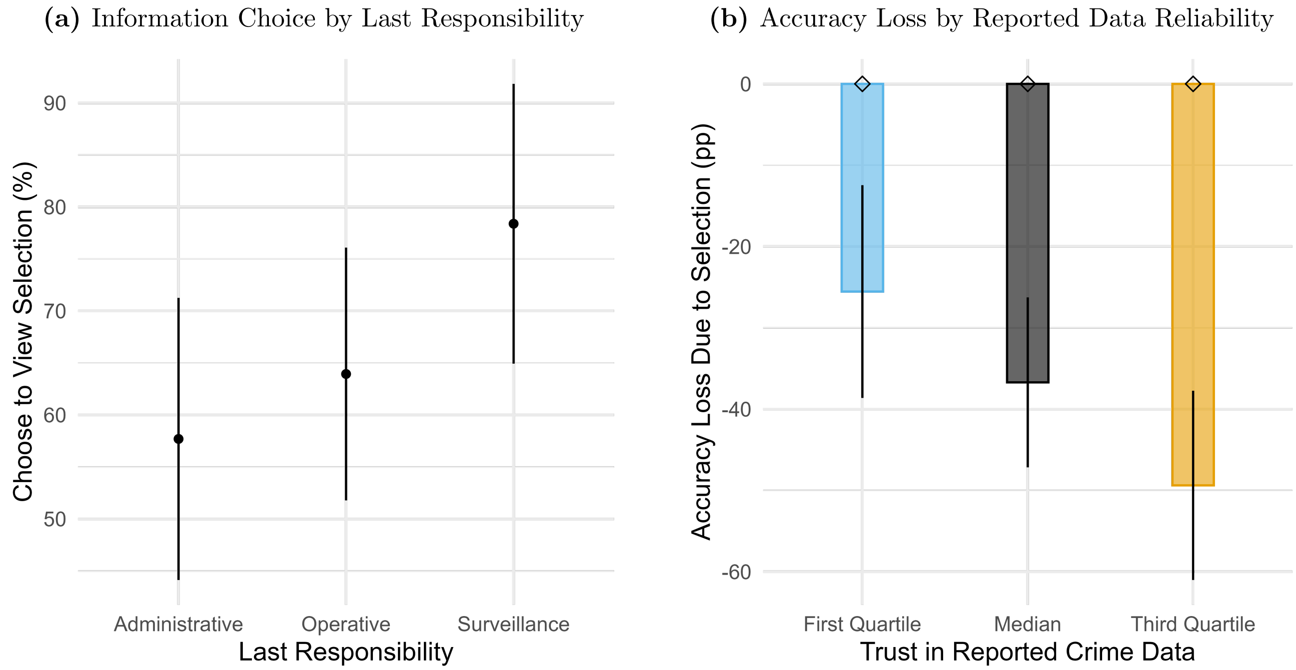
We begin by examining the extensive margin: do officers include reporting rates in their inference process at all? Before the tenth prediction of Part I, officers choose which variable to view alongside the number of reports. Selecting the reporting rate indicates that they understand that crime reports are a selected sample from the underlying crime pool. Only 63% of officers choose to view the reporting rate, meaning that 37% opt to ignore selection even when the data generating process is transparent. This establishes that a non-trivial share of officers exclude data selection from their prediction process.

Why does selection fail to come to mind for so many officers? If neglect were primarily driven by bottom-up visual salience, i.e., by perceptual cues, then making reporting rates difficult to ignore should reduce neglect. If instead selection neglect reflects an incorrect mental model, then salience manipulations should have limited impact. To distinguish these mechanisms, we first hold salience fixed and exploit variation in officers' beliefs about the reliability of the crime reports they routinely work with. At the end of the experiment, officers report their agreement with statements about whether crime reports accurately capture spatial and temporal crime patterns, whether citizens know how and where to report crimes, and whether reports provide a biased picture of crime. We combine these measures into a metric that we denote *reported data reliability*[7], where higher values reflect a stronger belief that reported crime data reliably represents underlying crime.

Figure 3 shows that officers who hold strong beliefs about the accuracy of reported crime are substantially less likely to incorporate data selection rates when forming crime predictions, while officers who view real crime reports as incomplete or biased are much more likely to account for selection. Officers at the third quartile of perceived reliability lose 23.8pp more prediction accuracy than those at the first quartile. Table A.4 confirms the heterogeneous effect of selection by perceived reliability of reported crime data (interaction $p < 0.001$). Consistent with selection neglect coming from established mental models, officers whose last operational responsibility involved surveillance operations (predicting crime and assigning patrols) are 20.7 pp ($p = 0.035$) more likely to choose to view the reporting rate than administrative officers.

[7] We follow Kling et al. (2007) and aggregate z-scores of each component of the index.

Figure 3: Selection Neglect Heterogeneity



Notes: Panel (a) shows the fraction of officers who choose to view selection for the last decision of Part I, disaggregated by their last responsibility. Error bars show the 95% confidence interval of the means. Panel (b) shows the accuracy loss due to selection, calculated as the marginal effect of selection-sensitivity on accuracy in Part II, disaggregated by reported data reliability. We plot the marginal effect of selection-sensitivity for 3 values of reliability: the first quartile (-0.27), median (0.01), and third quartile (0.34). Error bars show the 95% confidence interval of the marginal effect. Diamonds show optimal benchmark.

These patterns reinforce external validity —officers behave in the experiment according to their beliefs about real crime data— and are difficult to reconcile with a bottom-up salience interpretation. Whether selection enters crime predictions depends on officers' prior beliefs about how informative reported crime is, not on how prominently reporting rates are displayed.

The salience manipulation in Part II provides direct evidence that selection neglect isn't driven by perceptual salience. In Part II, officers only see crime data (number of reports) and its selection (reporting rates), making data selection more salient. Figure 4 finds that directing attention to reporting rates in this way doesn't improve prediction accuracy for those who previously ignored selection. Similarly, it doesn't alter much the distribution of accuracy for those who had chosen to view selection. Making selection visually unavoidable makes it more likely to be seen, but does not generate understanding of the adjustment rule. In short, bottom-up salience can increase exposure to information but does not substitute for the belief that selection must be corrected and how.

Having established the extensive margin of neglect, we now examine the intensive margin: conditional on recognizing that reporting rates matter, why do many officers still fail to adjust? Figure 4 provides a first indication. Among officers who choose to attend to selection in Part I, 27% never adjust successfully and 55% do so only intermittently. This is not random noise: the very same officers make accurate predictions in 78% of selection-neutral decisions, where taking the number of reports at face

Figure 4: Accuracy Distributions in Selection-Sensitive Decisions

Notes: This figure shows the distribution of officer-level prediction accuracy in selection-sensitive decisions. Plots in the first row show predictions from Part I, where data selection is less salient. Plots in the second row show prediction from Part II, where data selection is made more salient. Columns separate the sample between officers who chose *not* to view the reporting rate for the last prediction of Part I (left, in black), and officers who chose to view selection (right, in blue). When comparing across columns, we notice the strong link between choosing to see selection and prediction accuracy in Part I, where selection wasn't salient. Once selection is made more salient in Part II, officers who hadn't chosen to view it in Part I (left column) don't really improve accuracy but become more likely to systematically neglect it. For officers who already were paying attention to reporting rates, the increase in salience doesn't seem to have any effect. In both cases, salience doesn't mitigate neglect. While bottom-up salience increases the likelihood that selection comes to mind, it does not substitute for having the correct adjustment rule.

value leads to the optimal prediction. Two additional results point to computational frictions. First, accurate adjustment in selection-sensitive predictions takes 0.5 standard deviations longer than inaccurate adjustment ($p < 0.001$), while response times do not differ in selection-neutral predictions. Second, Table A.2 documents that officers are 9.9pp ($p = 0.048$) more likely to guess accurately when reporting rates are simple multiples of each other, and this gain in accuracy comes mostly from selection-sensitive decisions. When the computation becomes trivial, the correct adjustment is more likely to be executed. Figure A.1 provides further evidence that ease of computation moves participants away from occasional mistakes into consistent adjustment. In other words, even among officers who know that selection must be corrected, the computation is not consistently performed due its cognitive costs.

A final piece of evidence reinforces the dual-margin mechanism. Sixteen percent of officers use the on-screen calculator at least once. Calculator use can only improve predictions if officers already understand that selection must be corrected and how to correct it. Consistent with this logic, Table A.2 finds that calculator users perform similarly to non-users in selection-neutral predictions but are significantly more accurate in selection-sensitive ones.[8] Among calculator users, the accuracy loss due to selection falls by 27.8 pp ($p = 0.008$) and becomes statistically insignificant at the decision level. Once officers (i) recognize

[8] If calculator use merely captured motivation or effort, users should outperform in both types of predictions; we find no such difference in selection-neutral predictions ($p = 0.811$).

that reporting rates must be incorporated and (ii) are relieved of computation costs, selection neglect essentially disappears.

Selection neglect in policing is not a single error. Whether officers adjust depends first on whether data selection enters their mental model of crime data (extensive margin) and then on whether the adjustment is executed when it does (intensive margin). This dual structure explains why some officers never adjust, why many adjust only intermittently, why adjustment depends on cognitive load, and why selection neglect disappears only when both understanding and computation are in place.

These findings from the officers who make crime predictions on the field deepen and extend laboratory evidence on selection neglect. Consistent with Enke (2020), the extensive margin plays a central role, but our results reveal that whether selection comes to mind depends on officers' beliefs about the reliability of real-world crime reports: a top-down component driven by field experience rather than salience alone. The bimodal structure of accuracy parallels the representational differences emphasized by Bordalo et al. (2023), while the sensitivity of correct adjustment to computational complexity echoes laboratory evidence on costly reasoning (Arrieta & Nielsen, 2024; Farina & Herman, 2025). Together, the results provide a unified behavioral account of how officers process crime data and show that selection neglect in policing stems not only from statistical complexity, but from domain-specific mental models of crime data.

These results establish the prevalence of selection neglect in crime predictions and the mechanisms behind it. We now examine its consequences for disparities in inference.

V The Consequences of Selection Neglect

A Neglected crimes due to selection

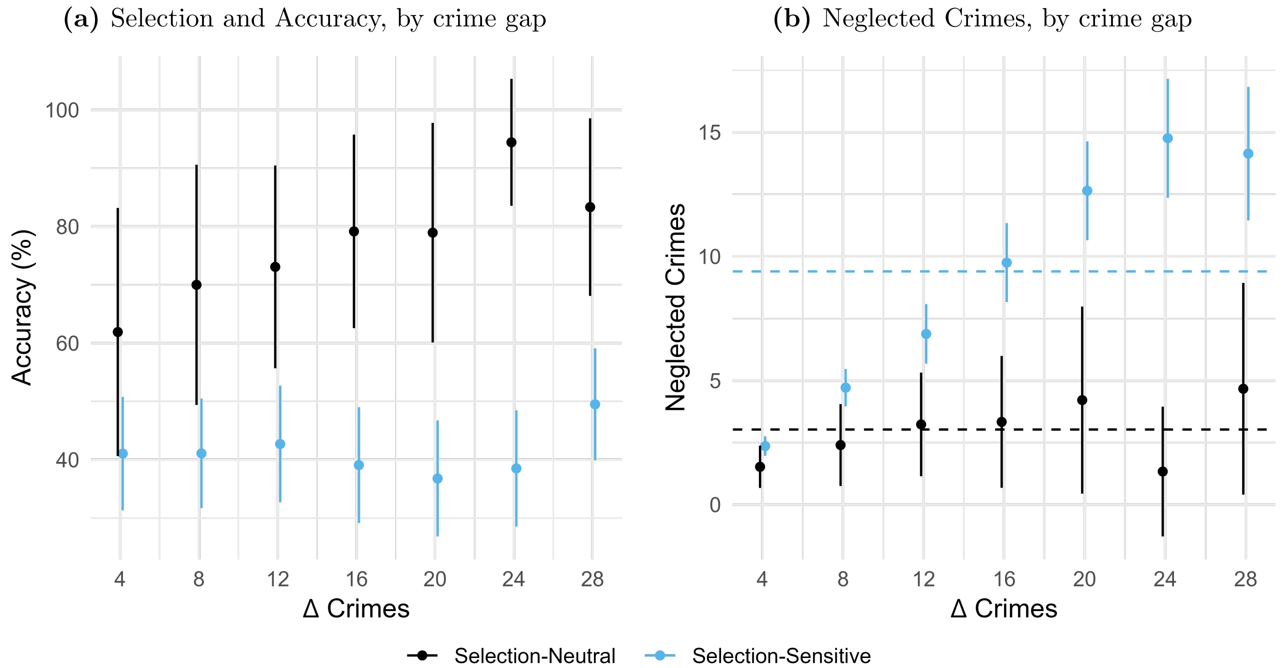
We now quantify the consequences of selection neglect for crime prediction. Because the optimal prediction is to identify the neighborhood with more crimes, we can compute how many crimes officers *neglect* relative to the optimal prediction: $c_{max} - c_{chosen}$. Neglected crimes translate the accuracy loss documented above into a metric directly relevant for policing: how much crime goes unaccounted for because of inference error. Although accuracy loss could in principle be operationally negligible if neglect occurred only in "close calls," i.e., cases with a small *crime gap*[9] between neighborhoods, the evidence below shows the opposite: selection neglect is most damaging precisely when the crime gap is large.

Panel (a) of Figure 5 shows that accuracy increases sharply with the crime gap in selection-neutral decisions. This is expected: when correcting for selection does not change the optimal prediction, a large crime gap also implies a large gap in reported crimes, making the choice easier. In contrast, accuracy remains flat in selection-sensitive decisions. Because selection neglect is captured by the accuracy gap between these two types of decisions, net of shared task complexity, selection neglect grows when the crime gap widens. This means officers are most likely to ignore data selection in precisely the situations where prediction errors carry the highest cost.

Panel (b) shows the implied cost in neglected crimes. Although the marginal cost of a mistake rises mechanically with the crime gap, the *realized* cost remains low in selection-neutral decisions: officers miss

[9] We define the crime gap as $\Delta Crimes_n = |c_{cn} - c_{dn}|$.

Figure 5: Neglected Crimes due to Selection in Part II



Notes: Panel (a) shows the average accuracy of patrolling decisions, by the absolute difference in crimes between neighborhoods. Error bars show the 95% confidence interval of the means, clustered at the officer level. Panel (b) shows the mean number of crimes neglected ($c_{max} - c_{chosen}$), by the difference in crimes between neighborhoods. Error bars show the 95% confidence interval of the means, clustered at the officer level, and dashed lines show the means for both types of decisions. Colors in both graphs show whether selection bias affects the optimal decision: it doesn't for *Selection-Neutral* (black), but it does for *Selection-Sensitive* decisions (blue). Both panels only use the first 6 decisions of Part II, where participants only see the number of reports and reporting rate.

about 3 crimes on average, and this number barely changes as the gap widens, since accuracy improves in parallel. In selection-sensitive decisions, however, neglected crimes increase steeply with the gap because accuracy does not improve. Across decisions, officers miss 6.4 additional crimes due to selection neglect (210%, $p < 0.001$). Given that the average crime gap between neighborhoods in Part II is 15.9 crimes, losses attributable solely to neglect represent roughly 40% of the average crime gap.

These results demonstrate that selection neglect is not merely a behaviorally relevant pattern: it produces materially worse crime predictions in the highest-stakes decisions. We next examine whether this mechanism can generate systematic disparities in crime predictions across groups.

B Selection Neglect and Disparities in Crime Predictions

In our experiment, the neighborhood guess serves purely as a discrete belief elicitation choice: officers choose the neighborhood they believe has more crime. When two neighborhoods have no crime gap in expectation, accurate predictions should be uniformly distributed across the two locations. However, if officers neglect selection, they may infer higher crime in over-selected neighborhoods, resulting in uneven crime predictions even when true crime is equal. To test this mechanism at the decision level, we regress crime predictions on the reporting-rate gap, conditional on the crime gap. Table 3 reports estimates from linear probability models. In Column (1), conditional on the true crime gap, officers become 0.4 pp more likely to predict higher crime in a neighborhood for every 1 pp higher reporting gap. Thus, positive selection bias in crime data translates directly into biased crime predictions. Column (2) shows that this effect is not driven only by officers for whom selection fails to come to mind: those who chose to view reporting rates appear, if anything, slightly more sensitive to reporting gaps, although the difference is not statistically significant. Columns (3) and (4) replicate these results in Part II, where reporting rates are made salient for all officers, supporting the same conclusion.

Table 3: Effect of crime and reporting gaps in prediction probability

	Part I - Choose A		Part II - Choose C	
	(1)	(2)	(3)	(4)
Crime Gap	0.003***	0.002*	0.003**	-0.0006
	(0.0005)	(0.0009)	(0.001)	(0.002)
Reporting Gap	0.004***	0.003***	0.003***	0.003***
	(0.0006)	(0.0009)	(0.0006)	(0.001)
View Selection × Crime Gap		0.002*		0.005**
		(0.001)		(0.002)
View Selection × Reporting Gap		0.0005		0.0005
		(0.001)		(0.001)
Fixed-effects				
Officer	Yes	Yes	Yes	Yes
Observations	1,224	1,224	1,236	1,236
Dependent variable mean	0.469	0.469	0.488	0.488
Part I Decoys	Yes	Yes	No	No

Notes: This table shows the results of regressing the neighborhood choice on the crime gap and the reporting gap between neighborhoods. Columns (1)-(2) use choices from Part I, so they control for the gaps in the demographic decoys (age, unemployment, and SES). Columns (3)-(4) use choices from Part II, where officers didn't see these decoys, so they are not included in the regressions. Columns (2) and (4) include an interaction with whether the officer chose to view selection (reporting rates) for the last prediction of Part I. All specifications include officer fixed effects. Standard errors in parentheses are robust and clustered at the officer level. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

The results above use random variation in reporting rates across decisions. In real policing environments, selection into crime data is typically not uniform: some groups are structurally more likely to be represented in crime data. For example, Black individuals are substantially more likely than White individuals to be arrested in Chicago and Seattle (Gonçalves et al., 2025); minority drivers are more likely to be stopped and cited by Florida Highway Patrol (Aggarwal et al., 2025); and low-income drivers are more likely to be searched by Texas Highway Patrol (Feigenberg & Miller, 2025). Even in systems driven by citizen reports, as in Colombia, systematic differences in who reports and which crimes get reported induce group-level selection bias (Perez-Vincent et al., 2024). If officers treat reported crime data as a faithful reflection of crime, selection neglect can convert structural over-selection into group-level disparities in crime predictions, even when true crime is identical across groups.

We test this prediction by exogenously manipulating group-level selection. In the *Systematic Selection Bias* treatment, one of the two cities (our abstract proxy for a "group") has a 30pp higher average reporting rate throughout Part II; in the *Uniform Selection Bias* control, reporting rates are equal in expectation. If selection neglect maps biased data into biased beliefs, officers in the systematic-selection treatment should systematically over-predict crime in the over-selected city.

Figure 6 confirms the prediction that systematic selection bias translates into disparities in predictions. Panel (a) shows that officers in the Systematic Selection Bias treatment predict higher crime in the over-selected city 18pp (43%) more often than in the under-selected city, despite both cities having identical true crime levels. No such asymmetry appears in the Uniform Selection Bias treatment. As a consequence, officers develop persistent prediction asymmetries: Panel (b) shows that only 5.5% of officers in the uniform-selection condition predict the same city to have more crime across all six decisions, a rate consistent with chance, while 20.6% do so in the systematic-selection condition. Table A.6 confirms these effects are statistically significant. These disparities emerge in an environment with no stereotypes, no taste-based bias, and no operational constraints: they arise purely from the interaction between biased data and selection neglect.

Finally, policing systems are not only characterized by structural selection: they also work with endogenously generated data. Predictions determine policing, policing determines what crime events are selected into crime data, and future predictions respond to the selected crime data. In theory, neglecting this endogeneity can sustain cycles of bias and over-policing (Che et al., 2024; Hübert & Little, 2023). We evaluate this dynamic by randomizing officers in the last six rounds of Part II into an *Endogenous Selection* treatment, where choosing a city increases its reporting rate by 15 pp in the next decision.

Selection neglect induces cycles of bias when selection is endogenous. Endogenous selection increases the probability of predicting the same city in any pair of consecutive decisions by 14.6 pp ($p = 0.027$), from a 53% baseline. Consequently, it increases the share of officers who consistently predict the same city to have more crime across six rounds. The effect is largest when systematic selection bias and endogenous selection coincide: in that setting, 37% of officers repeatedly over-predict crime in the over-selected city.

Taken together, these results show how selection neglect maps biased data into biased beliefs and how endogenous selection can reinforce those beliefs over time. In a simplified, high-information environment with no stereotypes, no taste-based preferences, and no operational constraints, selection neglect alone generates persistent and self-reinforcing disparities in crime predictions. This mechanism represents a distinct behavioral pathway to statistical discrimination: one that does not require animus, stereotypes, or differing marginal returns to policing.

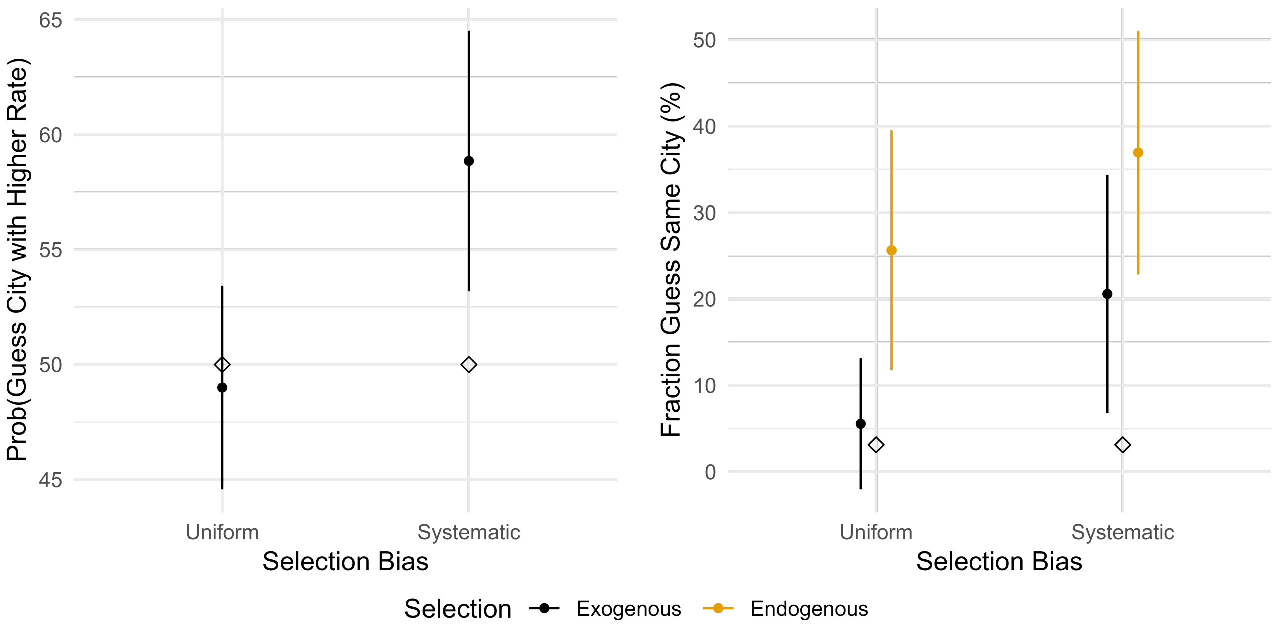
VI How to debias crime predictions

A Predictive Algorithms

Predictive algorithms have become increasingly common in policing agencies, motivated by the hope that statistical tools can correct systematic errors in human inference when predicting crime (Brayne,

Figure 6: Policing disparities driven by selection bias

(a) Choice of city with higher reporting rate **(b)** Fraction of sample who choose same city



Notes: Panel (a) shows the probability of guessing the city with the highest average reporting rate in Part II, by selection bias treatment. The left column shows officers in the *Uniform Selection Bias* treatment, where reporting rates for both cities are equal in expectation. The right column shows officers in the *Systematic Selection Bias*, where reporting rates are 30pp higher in expectation for one city. Error bars show the 95% confidence interval of the means, with standard errors clustered by officer. The diamonds show the optimal benchmark: because both cities have the same number of crimes in expectation, the optimal guesses are equally distributed. Panel (b) shows the fraction of officers who always guess the same city across the 6 last decisions of Part II, by the selection bias and endogeneity treatments. The left column shows officers in the *Uniform Selection Bias* treatment, while the right column shows officers in the *Systematic Selection Bias* treatment. Colors represent the endogeneity treatments. In black, we show officers in the *Exogenous Selection* control, where crime predictions don't affect reporting rates, and in orange the *Endogenous Selection* treatment, where guessing one city increases reporting rates by 15pp in that same city for the next decision. Error bars show the 95% confidence interval of the means, with standard errors clustered by officer. The diamonds show the optimal benchmark: because crimes are equally distributed across cities, the chance that it's optimal to guess the same city in 6 consecutive decisions is $2 \times 0.5^6 = 3.1\%$.

2020). When properly designed[10], these systems can reduce costly prediction errors. Yet in virtually all deployments, algorithms provide recommendations rather than making autonomous decisions, leaving discretion to human officers who may accept or override the algorithmic recommendations. The underlying assumption is that algorithms aggregate observable information while officers' experience captures private, localized knowledge. However, this division of labor has an important failure mode: if officers override recommendations exactly when those recommendations *correct* their inference errors, algorithmic tools provide little benefit. Next, we test whether algorithmic recommendations compensate for officers' inference errors or whether selection neglect extends to ignoring the algorithm itself.

Officers were randomly assigned to one of three treatments for the last six decisions of Part II. Officers in the *None* condition continued making decisions as before. Officers in the *Correcting Algorithm* and

[10] For example, recent research documents that predictive policing algorithms such as PredPol can amplify racial disparities because they themselves neglect endogenous selection in crime data (Arnold et al., 2025; Lum & Isaac, 2016).

Neglecting Algorithm treatments were informed that an algorithm would predict which neighborhood had more crime, and that they could either follow or override its recommendation. Crucially, the two algorithms differed only in whether they *adjusted* for data selection. Those in the *Correcting Algorithm* treatment were explicitly told that the algorithm accounted for reporting rates; those in the *Neglecting Algorithm* treatment were told that the algorithm relied only on raw reports. Both algorithms predicted the higher-crime neighborhood based on their respective inference rule, with some noise [11]. Therefore, following either algorithm should improve prediction accuracy in selection-neutral decisions, while the *Correcting* algorithm should improve and the *Neglecting* algorithm should worsen accuracy in selection-sensitive decisions.

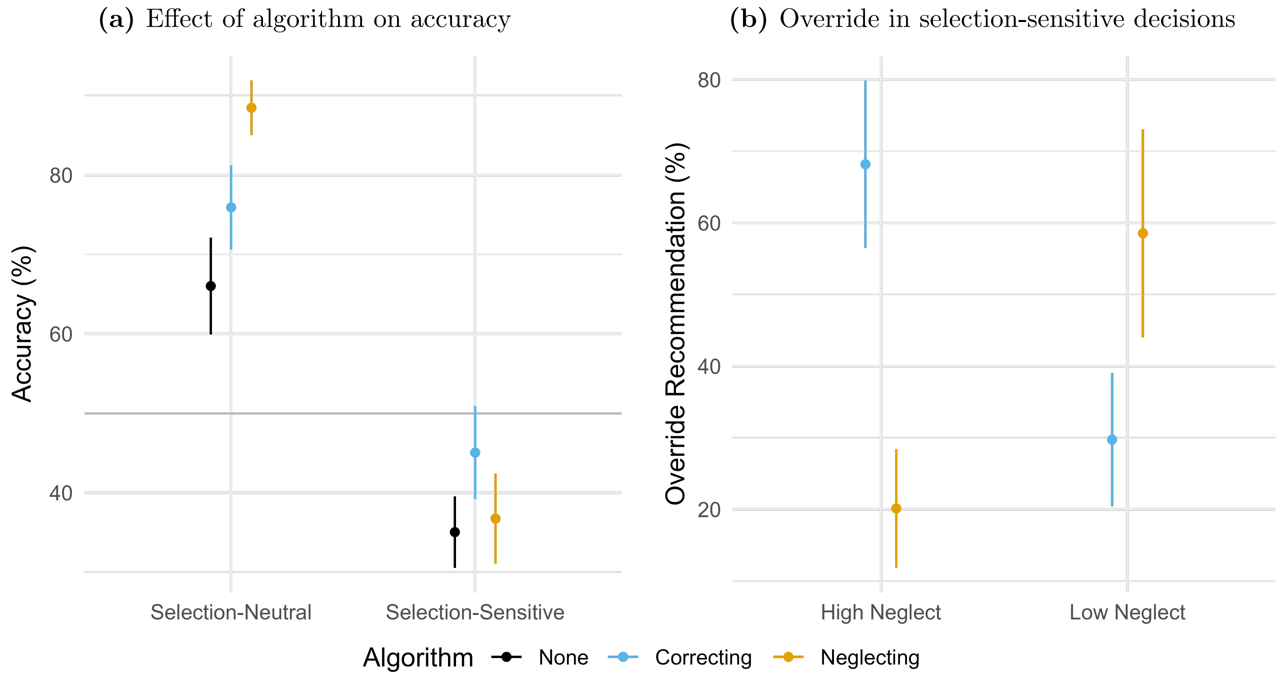
This design provides a diagnostic test of the mechanism driving selection neglect. For officers neglecting selection at the intensive margin, the *Correcting* algorithm should act as a "computation substitute" and they should follow it, especially in selection-sensitive decisions. In contrast, if selection simply does not come to mind, officers may override the *Correcting* algorithm because its recommendation contradicts their neglecting inference rule (that more reports imply more crime), while following the *Neglecting* algorithm because it reinforces that rule. Thus, algorithms are not only a potential remedy for selection neglect: they also reveal whether neglect originates from computational frictions or from a stable mental model of crime data.

Panel (a) of Figure 7 finds that algorithms do *not* reduce selection neglect. Both algorithms improve accuracy in selection-neutral decisions, but neither significantly increases accuracy in decisions where adjusting for selection is critical. Why does an algorithm that literally points to the correct choice not improve performance? Panel (b) suggests the answer: selection neglect extends to ignoring algorithmic recommendations. We divide the sample in half, above and below the median accuracy (33%) for selection-sensitive cases among the first six decisions of Part II (before the algorithms are introduced). Officers with high levels of selection neglect override the algorithm that corrects their heuristic in 68% of selection-sensitive decisions, yet override the neglecting algorithm only 20% of the time. Officers with low neglect show the opposite pattern: overriding the neglecting algorithm while following the correcting one. In other words, officers selectively trust algorithms when the recommendation *validates* their inference rule and override them when the recommendation *contradicts* it.

These results parallel findings from judicial settings in which judges override risk-prediction algorithms and frequently make lower-quality decisions as a result (Angelova et al., 2025). Taken together, the patterns of adherence and override support two conclusions. First, selection neglect is sufficiently entrenched that it affects not only independent inference but also responses to algorithmic support. Second, algorithms fail not because they are ignored indiscriminately, but because the very officers who neglect selection override the correcting algorithm exactly when it attempts to fix their mistake. Algorithms do not substitute for a missing inference rule: they are filtered through it.

[11] In practice, both algorithms made a correct prediction for over 90% of the selection-neutral decisions, while for selection-sensitive decisions the hit rate was 78% for the *Correcting* algorithm and 13% for the *Neglecting* one.

Figure 7: Algorithmic recommendations: adherence and effect on accuracy



Notes: shows the average prediction accuracy in selection-neutral decisions (left column) and selection-sensitive decisions (right column). Panel (b) shows the probability of overriding a recommendation in selection-sensitive decisions, dividing the sample in half according to their level of neglect in the first 6 decisions of Part II. The left column shows officers with *High Neglect* —below-median accuracy ($< 33\%$) in selection-sensitive decisions, while the right column shows officers with *Low Neglect* —above the median. In both panels, error bars show the 95% confidence interval of the means, clustered at the officer level. Colors show the algorithm treatments: *None* in black, *Correcting* in blue, and *Neglecting* in orange.Colors represent the same algorithm interventions.

VII Discussion

This paper shows that many police officers neglect the selection of crime data when forming beliefs about crime, even when the data-generating process and its selection are transparent. In the presence of large and transparent selection bias, a large share of officers treat reported crime data as if it was representative of underlying crime. The evidence indicates that neglect is not a single cognitive mistake but the joint result of two frictions: for some officers, the selection problem does not enter their mental model of crime data, while for others it does enter but the required adjustment is applied only when it is computationally easy. These frictions are systematically related to officers' beliefs about the reliability of reported crime data in the field, implying that real-world experience facilitates adjusting for selection. Because selection neglect is strongest precisely when crime differences are large, it has first-order consequences: biased data becomes biased beliefs, and, when data selection responds to policing, biased beliefs feed back into more biased data. Our results thus document that differences in crime observability across groups translate directly into group disparities in crime beliefs among police officers, even in the absence of taste bias, stereotypes, or institutional pressures.

Our findings build on and advance an active literature on how people approach statistical problems in general (Bohren et al., 2024; Bordalo et al., 2023), and selection problems in particular (Enke, 2020). A recent line of experimental work has shown that experimental subjects often fail to adjust for selection

(Ali et al., 2021; Araujo et al., 2021; Barron et al., 2024; Esponda & Vespa, 2018; Farina & Herman, 2025; Jin et al., 2021). Our paper advances this literature in two ways. First, it provides evidence that selection neglect is pervasive not only among experimental subjects doing an abstract task, but also among police officers when doing a familiar crime prediction task. Second, our paper doesn't stop at documenting the extent and consequences of selection neglect, but carefully identifies the mechanisms behind it and connects them with officers' prior experience in the field and their awareness of selection bias in real crime data. Our unique sample of patrol-assigning police officers offers the perfect ground to test the external validity of lab findings among high-stakes decision makers. Thus, our paper is close to other work focusing on what drives these failures in contingent thinking (Enke, 2020; Martínez-Marquina et al., 2019), and on whether experts in the field are immune to them (Barrios et al., 2025; Koehler & Mercer, 2009; Malmendier & Shanthikumar, 2007).

A related literature analyzes the cognitive processes of decision makers in the criminal justice system. Using administrative data on sentencing, bail, pretrial release, or stop and search decisions, a nascent literature explores how agents in the criminal justice system form and update beliefs (Angelova et al., 2025; Arnold et al., 2018; Bhuller & Sigstad, 2024; Feigenberg & Miller, 2025; Prendergast, 2021). Our paper contributes to this recent work, but instead of inferring beliefs from decisions, elicits crime predictions directly. This allows us to isolate the cognitive mechanism and abstract from operational constraints, while remaining agnostic about the exact mapping from crime predictions to policing decisions. Additionally, we provide evidence that an unbiased predictive algorithm doesn't improve crime predictions, because officers who neglect selection override the algorithm's corrective recommendations, consistent with previous findings from judges (Angelova et al., 2025) and doctors (Mullainathan & Obermeyer, 2022).

Finally, this paper documents a cognitive mechanism that might amplify discrimination in policing. Once the racial and class disparities in policing have been established (National Academies of Sciences, Engineering, and Medicine, 2023), recent work has aimed to identify their sources, focusing on the taste bias of individual officers (Ba et al., 2021; Goncalves & Mello, 2021; Hoekstra & Sloan, 2022; Rozema & Schanzenbach, 2019), and institutional constraints (Ba et al., 2025; Ferrazares, 2024; Rozema & Schanzenbach, 2023). A promising line of work focuses on the cognitive roots of discrimination, providing evidence that cognitive depletion (Ferrazares, 2025), emotional responses (Holz et al., 2023), and impulsive thinking (Dube et al., 2024; Owens et al., 2018), are related to excessive use of force and racial disparities. Our paper takes a similar stand, highlighting that racial differences in crime observability lead to systematic biases in crime beliefs across groups. While Gonçalves et al. (2025) document selection bias in crime data, and Arnold et al. (2025) show that predictive algorithms replicate this bias in their recommendations, we show that selection neglect among police officers maps selection bias into biased crime beliefs. Although crime predictions are a crucial input in policing decisions (Brayne, 2020), focuses on isolating the belief formation mechanism without making claims about how biased beliefs map into policing decisions.

Taken together, our findings show that unequal observability does not merely distort crime datasets, it distorts crime beliefs. When officers neglect the selection of crime data, selection bias becomes prediction bias, and prediction bias feeds back into future selection when policing affects observability. This mechanism is distinct from, and complementary to, taste-based, stereotype-based, and institutional explanations of disparities. While we take no stance on the relative empirical importance of each mechanism, our evidence shows that belief formation represents a critical bottleneck in policing decisions: without correcting for selection bias, even unbiased officers or well-intended algorithms can generate persistent disparities in crime predictions. Future research can build on these insights by measuring how biased crime beliefs translate into real policing decisions and by developing interventions that target not only data quality but the cognitive processes through which officers interpret the data.

References

Aggarwal, P., Brandon, A., Goldszmidt, A., Holz, J., List, J. A., Muir, I., Sun, G., & Yu, T. (2025). High-frequency location data show that race affects citations and fines for speeding. *Science*, *387*(6741), 1397–1401.

Ali, S. N., Mihm, M., Siga, L., & Tergiman, C. (2021). Adverse and Advantageous Selection in the Laboratory. *American Economic Review*, *111*(7), 2152–2178.

Ang, D., Bencsik, P., Bruhn, J., & Derenoncourt, E. (2025). Community Engagement with Law Enforcement after High-Profile Acts of Police Violence. *American Economic Review: Insights*, *7*(1), 124–142.

Angelova, V., Dobbie, W., & Yang, C. S. (2025). Algorithmic Recommendations and Human Discretion. *The Review of Economic Studies*, rdaf084.

Araujo, F. A., Wang, S. W., & Wilson, A. J. (2021). The Times They Are A-Changing: Experimenting with Dynamic Adverse Selection. *American Economic Journal: Microeconomics*, *13*(4), 1–22.

Arnold, D., Dobbie, W., & Hull, P. (2025). Building Nondiscriminatory Algorithms in Selected Data. *American Economic Review: Insights*, *7*(2), 231–249.

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932.

Arrieta, G., & Nielsen, K. (2024). Procedural Decision-Making In The Face Of Complexity.

Ba, B., Chen, P., Cheng, T., Eies, M. C., & Holz, J. E. (2025, October). *What is the Best Response? Examining the Impact of Police and Their Alternatives* (tech. rep. No. w34344). National Bureau of Economic Research.

Ba, B., Knox, D., Mummolo, J., & Rivera, R. (2021). The role of officer race and gender in police-civilian interactions in Chicago. *Science*, *371*(6530), 696–702.

Barrios, J. M., Lancieri, F., Levy, J., Singh, S., Valletti, T., & Zingales, L. (2025, April). *The Conflict-of-Interest Discount in the Marketplace of Ideas* (tech. rep. No. w33645). National Bureau of Economic Research.

Barron, K., Huck, S., & Jehiel, P. (2024). Everyday Econometricians: Selection Neglect and Overoptimism When Learning from Others. *American Economic Journal: Microeconomics*, *16*(3), 162–198.

Bhuller, M., & Sigstad, H. (2024). Feedback and Learning: The Causal Effects of Reversals on Judicial Decision-Making. *The Review of Economic Studies*, rdae073.

Bohren, J. A., Hascher, J., Imas, A., Ungeheuer, M., & Weber, M. (2024, January). A Cognitive Foundation for Perceiving Uncertainty.

Bohren, J. A., & Hauser, D. N. (2025). Misspecified Models in Learning and Games.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2023). How People Use Statistics. *NBER Working Paper*.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2022). Salience. *Annual Review of Economics*, *14*(Volume 14, 2022), 521–544.

Brayne, S. (2020, November). *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press.

Brundage, M., Little, A. T., & You, S. (2024). Selection Neglect and Political Beliefs. *Annual Review of Political Science*, *27*(Volume 27, 2024), 63–85.

Che, Y.-K., Kim, J., & Mierendorff, K. (2024, September). Predictive Enforcement.

Chen, M. K., Christensen, K. L., John, E., Owens, E., & Zhuo, Y. (2023). Smartphone Data Reveal Neighborhood-Level Racial Disparities in Police Presence. *The Review of Economics and Statistics*, 1–29.

DANE. (2023). *Encuesta de convivencia y seguridad ciudadana (ECSC)* (tech. rep.).

Dube, O., MacArthur, S. J., & Shah, A. K. (2024). A Cognitive View of Policing. *The Quarterly Journal of Economics*.

Enke, B. (2020). What You See Is All There Is. *The Quarterly Journal of Economics*, *135*(3), 1363–1398.

Enke, B., & Graeber, T. (2023). Cognitive Uncertainty. *The Quarterly Journal of Economics*, *138*(4), 2021–2067.

Esponda, I., & Vespa, E. (2018). Endogenous sample selection: A laboratory study. *Quantitative Economics*, *9*(1), 183–216.

Farina, A., & Herman, C. (2025). Communicating with Data Generating Processes: An Experimental Analysis.

Feigenberg, B., & Miller, C. (2025, March). *Class Disparities and Discrimination in Traffic Stops and Searches* (Working Paper No. 33629). National Bureau of Economic Research.

Ferrazares, T. (2024). Monitoring Police with Body-Worn Cameras: Evidence from Chicago. *Journal of Urban Economics*, *141*, 103539.

Ferrazares, T. (2025). Shift Structure and Cognitive Depletion: Evidence from Police Officers.

Fryer, R. G. (2019). An Empirical Analysis of Racial Differences in Police Use of Force. *Journal of Political Economy*, *127*(3), 1210–1261.

Goncalves, F., & Mello, S. (2021). A Few Bad Apples? Racial Bias in Policing. *American Economic Review*, *111*(5), 1406–1441.

Gonçalves, F. M., Mello, S., & Weisburst, E. K. (2025, August). *Selection Bias and Racial Disparities in Police Use of Force* (tech. rep. No. w34175). National Bureau of Economic Research.

Hoekstra, M., & Sloan, C. (2022). Does Race Matter for Police Use of Force? Evidence from 911 Calls. *American Economic Review*, *112*(3), 827–860.

Holz, J. E., Rivera, R. G., & Ba, B. (2023). Peer Effects in Police Use of Force. *American Economic Journal: Economic Policy*, *15*(2), 256–291.

Hübert, R., & Little, A. T. (2023). A Behavioural Theory of Discrimination in Policing. *The Economic Journal*, *133*(655), 2828–2843.

Jehiel, P. (2018). Investment Strategy and Selection Bias: An Equilibrium Perspective on Overoptimism. *American Economic Review*, *108*(6), 1582–1597.

Jin, G. Z., Luca, M., & Martin, D. (2021). Is no news (perceived as) bad news? An experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, *13*(2), 141–173.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental Analysis of Neighborhood Effects. *Econometrica*, *75*(1), 83–119.

Koehler, J. J., & Mercer, M. (2009). Selection Neglect in Mutual Fund Advertisements. *Management Science*, *55*(7), 1107–1121.

López-Pérez, R., Pintér, Á., & Sánchez-Mangas, R. (2022). Some conditions (not) affecting selection neglect: Evidence from the lab. *Journal of Economic Behavior & Organization*, *195*, 140–157.

Lum, K., & Isaac, W. (2016). To Predict and Serve? *Significance*, *13*(5), 14–19.

Malmendier, U., & Shanthikumar, D. (2007). Are small investors naive about incentives? *Journal of Financial Economics*, *85*(2), 457–489.

Martínez-Marquina, A., Niederle, M., & Vespa, E. (2019). Failures in Contingent Reasoning: The Role of Uncertainty. *American Economic Review*, *109*(10), 3437–3474.

Mullainathan, S., & Obermeyer, Z. (2022). Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care. *The Quarterly Journal of Economics*, *137*(2), 679–727.

National Academies of Sciences, Engineering, and Medicine. (2023, May). *Reducing Racial Inequality in Crime and Justice: Science, Practice, and Policy.* National Academies Press.

Owens, E., Weisburd, D., Amendola, K. L., & Alpert, G. P. (2018). Can You Build a Better Cop? *Criminology & Public Policy*, *17*(1), 41–87.

Perez-Vincent, S. M., Abril, V., Chen, C., Tayo, T., & Jimenez, A. U. (2024, October). *Crime Underreporting in Latin America and the Caribbean* (tech. rep.).

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745.

Prendergast, C. (2021, February). 'Drive and Wave': The Response to LAPD Police Reforms After Rampart.

Rivera, R. (2025). Do Peers Matter in the Police Academy? *American Economic Journal: Applied Economics*, *17*(2), 127–164.

Rozema, K., & Schanzenbach, M. (2019). Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct. *American Economic Journal: Economic Policy*, *11*(2), 225–268.

Rozema, K., & Schanzenbach, M. (2023). Does Discipline Decrease Police Misconduct? Evidence from Chicago Civilian Allegations. *American Economic Journal: Applied Economics*, *15*(3), 80–116.

Schwartzstein, J. (2014). Selective Attention and Learning. *Journal of the European Economic Association*, *12*(6), 1423–1452.

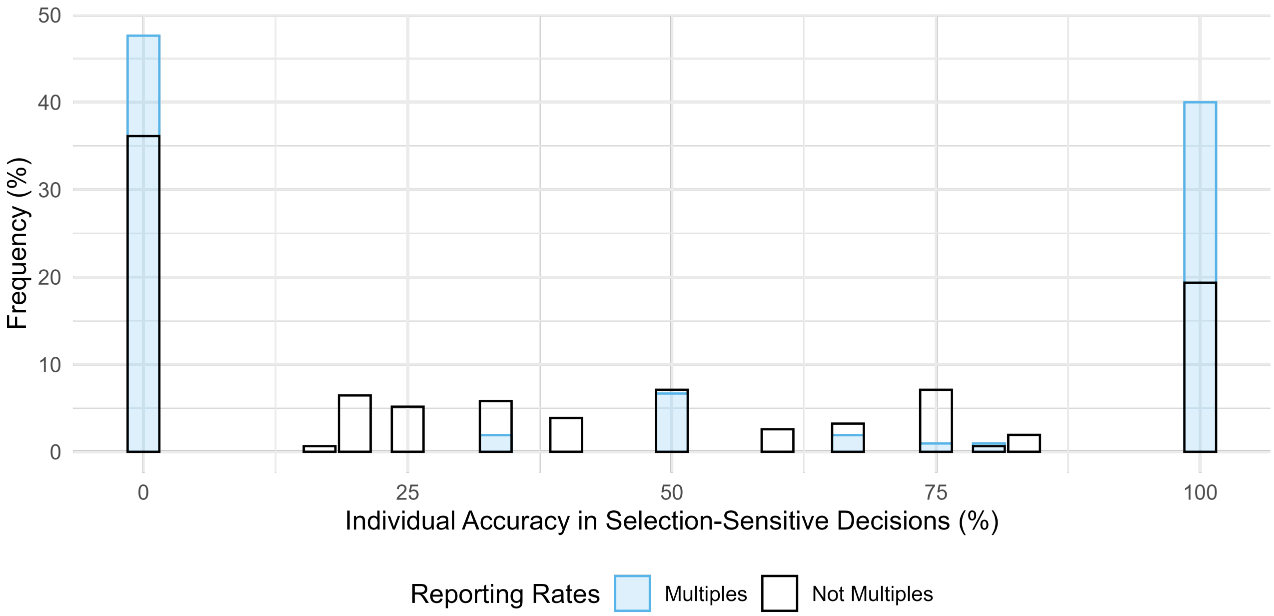
Stashko, A. (2023). Do Police Maximize Arrests or Minimize Crime? Evidence from Racial Profiling in U.S. Cities. *Journal of the European Economic Association*, *21*(1), 167–214.

Vomfell, L., & Stewart, N. (2021). Officer bias, over-patrolling and ethnic disparities in stop and search. *Nature Human Behaviour*, *5*(5), 566–575.

A Appendix

Additional Figures and Tables

Figure A.1: Accuracy in selection-sensitive decisions, by whether rates are multiples



Notes: This figure shows the distribution of officer-level prediction accuracy in selection-sensitive decisions in the first six decisions of Part II. We divide these decisions between those where reporting rates of the two neighborhoods in a decision are multiples of each other (blue), and those where the rates are not multiples (black). The figure documents that ease of calculation moves participants from occasional errors to consistent answers, with the fraction of officers always responding correctly more than doubling.

Table A.1: Descriptive Statistics

Sample Size	182
Male (%)	86.3
Age	36.3
Masters Degree (%)	15.4
Years in National Police	14.0
Last Task: Operative (%)	40.2
Last Task: Surveillance (%)	25.1
Instruction Errors (out of 9)	0.9
Math Errors (out of 2)	0.4
Pass Sanity Check (%)	85.2

Notes: Values are sample means unless otherwise specified

Table A.2: Prediction accuracy across decisions

	Part I			Part II		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.779***	0.673***	0.828***	0.770***	0.752***	0.775***
	(0.028)	(0.052)	(0.038)	(0.043)	(0.050)	(0.047)
S-Sensitive	-0.358***	-0.282***	-0.407***	-0.362***	-0.362***	-0.423***
	(0.042)	(0.071)	(0.060)	(0.057)	(0.063)	(0.062)
View Selection		0.165***				
		(0.060)				
S-Sensitive × View Selection		-0.119				
		(0.089)				
Rates are Multiples			0.032		0.051	
			(0.055)		(0.063)	
S-Sensitive × Rates are Multiples			0.068		0.031	
			(0.075)		(0.076)	
Use Calculator						-0.036
						(0.101)
S-Sensitive × Use Calculator						0.392***
						(0.111)
Observations	1,240	1,224	784	930	930	930
Dependent variable mean	0.507	0.507	0.532	0.467	0.467	0.467

Notes: This table shows the results of regressing prediction accuracy (dummy) on whether adjusting for selection was needed to make the correct prediction (*Selection-Sensitive decisions*). Columns (1)-(3) use predictions from Part I, while Columns (4)-(6) use predictions from Part II, where selection is made more salient. Column (2) interacts selection-sensitiveness with whether the participant chose to view reporting rates in the last prediction of Part I (*View Selection*). Column (3) adds an interaction with whether reporting rates in the two neighborhoods are multiples of each other (*Rates are Multiples*), focusing on the subset of subjects who chose to view selection. Column (5) does the same for Part II, but as selection is already made salient to all subjects, it includes all participants. Column (6) adds an interaction with whether the on-screen calculator was used for that decision. Standard errors in parentheses are robust and clustered at the officer level. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.3: Heterogeneity in choice to view selection

	(1)	(2)
Constant	0.577^{***}	0.649^{***}
	(0.069)	(0.039)
Operative	0.062	
	(0.093)	
Surveillance	0.207^{**}	
	(0.097)	
Reports Reliability		0.107
		(0.085)
Observations	150	151
Dependent variable mean	0.653	0.649

Notes: This table shows the results of regressing the indicator of whether the participant chose to view selection for the last decision of Part I, on last responsibility and perceived reliability of reported data. Column (1) shows that police officers with surveillance responsibilities (those predicting crime and assigning patrols) incorporate data selection in their model more often than those with administrative responsibilities. Column (2) shows no effect of perceived reported data reliability on the choice to view selection. Standard errors in parentheses are robust and clustered at the officer level. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.4: Heterogeneity in accuracy loss due to selection

	Part I		Part II	
	(1)	(2)	(3)	(4)
Constant	0.788***	0.782***	0.806***	0.767***
	(0.045)	(0.027)	(0.044)	(0.040)
Selection-Sensitive	-0.428***	-0.358***	-0.431***	-0.362***
	(0.071)	(0.041)	(0.083)	(0.054)
Operative	0.002		-0.042	
	(0.063)		(0.057)	
Surveillance	-0.038		-0.062	
	(0.073)		(0.073)	
Selection-Sensitive × Operative	0.131		0.103	
	(0.095)		(0.102)	
Selection-Sensitive × Surveillance	0.082		0.085	
	(0.117)		(0.121)	
Reports Reliability		0.115*		0.236***
		(0.066)		(0.084)
Selection-Sensitive × Reports Reliability		-0.255**		-0.388***
		(0.098)		(0.112)
Observations	1,216	1,224	1,824	918
Dependent variable mean	0.512	0.511	0.502	0.465

Notes: This table shows the results of regressing prediction accuracy (dummy) on whether adjusting for selection was needed to make the correct prediction (*Selection-Sensitive* decisions). Columns (1)-(2) use predictions from Part I, while Columns (3)-(4) use predictions from Part II, where selection is made more salient. Columns (1) and (3) interact selection-sensitiveness with the officer's last responsibility, taking as reference those with administrative tasks. Columns (2) and (4) add an interaction with the perceived reliability of reported crime data in real life. Standard errors in parentheses are robust and clustered at the officer level. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.5: Neglected crimes across decisions

	Part I		Part II	
	(1)	(2)	(3)	(4)
Constant	9.88***	10.5***	3.03***	1.74**
	(1.28)	(2.30)	(0.686)	(0.814)
Selection-Sensitive	11.3***	-8.72***	6.37***	-0.997
	(1.88)	(2.62)	(0.926)	(0.930)
Crime Gap		-0.011		0.081
		(0.043)		(0.067)
Selection-Sensitive × Crime Gap		0.516***		0.459***
		(0.068)		(0.084)
Observations	1,240	1,240	930	930
Dependent variable mean	18.4	18.4	8.35	8.35

Notes: This table shows the results of regressing neglected crimes ($c_{max} - c_{chosen}$) on decision features. Columns (1)-(2) use predictions from Part I, while Columns (3)-(4) use predictions from Part II, where selection is made more salient. Columns (1) and (3) regress neglected crimes on selection-sensitiveness, i.e., whether adjusting for selection changes crime predictions. Columns (2) and (4) add an interaction with the absolute crime gap between neighborhoods in the decision. Standard errors in parentheses are robust and clustered at the officer level. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A.6: Treatment effect of systematic selection bias and endogenous selection

Dependent Variables:	Chose High Rate	Repeated Guess	Fixed Guess
	(1)	(2)	(3)
Constant	0.491***	0.532***	0.056
	(0.021)	(0.048)	(0.039)
Systematic Bias	0.098***		0.150*
	(0.035)		(0.080)
Endogenous		0.146**	0.201**
		(0.065)	(0.081)
Systematic Bias × Endogenous			-0.037
			(0.129)
Observations	1,086	330	155
Dependent variable mean	0.540	0.609	0.232

Notes: This table shows the effects of the two treatments of Part II on several outcomes. Column (1) shows the effect of the *Systematic Selection Bias* treatment on the probability of choosing the neighborhood from the city with the higher average reporting rate across Part II. In that treatment, neighborhoods from one city have a 30pp higher reporting rate in expectation, while in the control condition neighborhoods from both cities have the same reporting rate in expectation. That one city has a higher average rate in the control is a result of the random sampling of reporting rates, but this reporting gap is small on average (less than 5pp). The effect is estimated at the decision level. Because both cities have the same underlying crime and reporting rates are independent of crime, officers adjusting for selection would guess the city with higher reporting rate half of the time in each treatment. Column (2) regresses the probability of guessing the same city in two consecutive predictions on the *Endogenous Selection* treatment, where choosing one city increases reporting rates in that city by 15pp in the next decision. This effect is estimated at the decision level, for the last six decisions of Part II and the officers who are randomized into not seeing algorithmic recommendations. Because both cities have the same underlying crime, officers adjusting for selection should repeat the same guess with 50% probability. Column (3) regresses the probability of always guessing the same city across the last six decisions of Part II (*Fixed Guess*) on the two treatments. The probability of a fixed guess being optimal is 3.1% for an officer correctly adjusting for selection. This effect is estimated at the officer level. Standard errors in parentheses are robust and clustered at the officer level. Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table A.7: Standardized response times by decision features

	Part I			Part II		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.024	0.404***	0.320**	0.029	0.472**	0.228
	(0.065)	(0.138)	(0.129)	(0.078)	(0.213)	(0.188)
Accurate	0.179**	-0.438***	-0.513***	0.263***	-0.503**	-0.466**
	(0.082)	(0.147)	(0.135)	(0.090)	(0.223)	(0.200)
Selection-Sensitive		-0.426***	-0.389***		-0.477**	-0.341*
		(0.142)	(0.130)		(0.220)	(0.194)
Selection-Sensitive × Accurate		0.802***	0.760***		0.919***	0.609***
		(0.176)	(0.159)		(0.260)	(0.231)
Use Calculator			1.86***			1.42***
			(0.123)			(0.143)
Observations	1,240	1,240	1,240	930	930	930
Dependent variable mean	0.115	0.115	0.115	0.151	0.151	0.151

Notes: This table shows the estimation of regressing standardized response times on some decision features. Columns (1)-(3) use predictions from Part I, while Columns (4)-(6) use the first six predictions from Part II, where selection is made more salient. Columns (1) and (4) regress the z-score of response time on whether the decision was accurate. Columns (2) and (5) interact this effect with whether the decision was selection-sensitive. Columns (3) and (6) add a control for the use of the on-screen calculator. Across specifications, we find that making the accurate adjustment in selection-sensitive decisions is time consuming. Standard errors in parentheses are robust and clustered at the officer level. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.